AMENDMENTS TO THE CLAIMS:

This listing of the claims will replace all prior versions, and listings, of the claims in this application.

Listing of Claims:

1. (Currently Amended) A method to process a text document, comprising:

partitioning text of the text document and assigning semantic meaning to words of the partitioned text, where assigning comprises applying a plurality of regular expressions, rules and ~~a plurality of~~ dictionaries comprising a common chemical prefix dictionary and a common chemical suffix dictionary to recognize chemical name fragments;

recognizing any substructures present in the chemical name fragments;

extracting keywords associated with the recognized chemical name fragments and the substructures of the text document and indexing the extracted keywords in a text index;

adding each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

determining structural connectivity information of each of the recognized chemical name fragments and ~~recognized~~ the substructures that do not contain a number;

~~extracting information associated with the recognized chemical name fragments and substructures of the text document and indexing the extracted information in a text index;~~

indexing representations of the recognized chemical name fragments and the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index;

storing the text index in association with the ~~indexed representations~~ chemical substructure index ~~in a searchable index~~; ~~and~~

providing a graphical user interface to search the ~~searchable~~ text index and the chemical substructure index, where the search comprises first entering search terms comprising one or more chemical fragment names and ~~entering~~ then selecting graphical representations of one or more substructures, ~~in a representation form, where the entering is by at least one of text form or graphical selection~~ where the selecting comprises using the graphical user interface as a pointer to a graphical list of substructures; and

receiving a search result, where the search result is an intersection of the chemical substructure index and the text index, identifying at least one document where there are found chemical compounds that contain the selected substructures, and connectivity specified by the one or more chemical fragment names and the selected substructures.

2. (Currently Amended) ~~A~~ The method as in claim 1, ~~wherein the extracting further comprises extracting keywords from the text document and indexing the keywords in the text index, and~~ wherein the search further comprises ~~selecting a graphical representation of one or more substructures and additionally~~ first entering search terms comprising the one or more chemical fragment names and entering at least one keyword, and where the search result is identifying at least one document where there are found the at least one keyword, the chemical compounds that contain the selected substructures, and the connectivity specified by the one or more chemical fragment names and the selected substructures.

3. (Currently Amended) A method as in claim 1~~, wherein extracting further comprises extracting keywords from the text document and indexing the keywords in the text index, and wherein the search comprises additionally entering at least one keyword, and at least one of chemical name fragment connectivity and substructure connectivity~~ performed by executing a computer program product.

4.– 6. (Cancelled)

7. (Currently Amended) ~~A~~ The method as in claim 1, where determining structural connectivity information comprises looking up recognized chemical name fragments and substructures in a structure dictionary.

8. (Currently Amended) ~~A~~ The method as in claim 1, where the indexing representations

~~comprise MOL type representations and SMILES type representations~~ of the recognized chemical name fragments and the substructures comprises:

testing if each of the recognized chemical name fragments occur in a SMILES fragment dictionary, where if it does occur in the SMILES fragment dictionary then adding the chemical name fragment to the chemical substructure index as the SMILES representation, and

testing if each of the recognized chemical name fragments occur in a MOL file fragment dictionary, where if it does occur in the MOL file dictionary then adding the chemical name fragments to the chemical substructure index as the MOL file representation.

9. (Currently Amended) ~~A~~ The method as in claim 1, where said plurality of dictionaries ~~comprise a~~ consists of the dictionary of common chemical prefixes and ~~a~~ the dictionary of common chemical suffixes.

10. (Currently Amended) ~~A~~ The method as in claim 1, where said plurality of dictionaries ~~comprise~~ consists of the common chemical prefix dictionary and the common chemical suffix dictionary, and a dictionary of stop words to eliminate erroneous chemical name fragments.

11. (Currently Amended) ~~A~~ The method as in claim 1, further comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

12. (Currently Amended) ~~A~~ The method as in claim 1, where chemical name fragments are

further recognized by using common chemical word endings.

13. (Currently Amended) ~~A~~ The method as in claim 1, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed from between chemical name fragments as a function of context.

14. (Currently Amended) ~~A~~ The method as in claim 1, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

15. (Currently Amended) ~~A~~ The method as in claim 14, where the punctuation comprises at least one of a parenthesis, a square bracket, a hyphen, a colon and a semi-colon.

16. (Currently Amended) ~~A~~ The method as in claim 14, where the characters comprise ~~at least one of~~ upper case C, O, R, N and H.

17. (Currently Amended) ~~A~~ The method as in claim 14, where the characters comprise ~~strings of at least one of~~ lower case xy, ene, ine, yl, ane and oic.

18. (Currently Amended) ~~A~~ The method as in claim 1, comprising an initial step of tokenizing the document to provide a sequence of tokens.

19. (Currently Amended) A system ~~to process a text document~~ having at least one computer, comprising:

a ~~unit~~ tokenizer module and a token processing module configured to partition text of the text document and to assign semantic meaning to words of the partitioned text~~, where assigning comprises~~ by applying a plurality of regular expressions, rules and ~~a plurality of~~ dictionaries comprising a common chemical prefix dictionary and a common chemical suffix dictionary to recognize chemical name fragments;

~~a unit~~ the token processing module configured to recognize any substructures present in the chemical name fragments;

the token processing module configured to extract keywords associated with the recognized chemical name fragments and the substructures of the text document and to index the extracted keywords in a text index;

the token processing module configured to add each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

~~a unit to extract information associated with the recognized chemical name fragments and substructures of the text document and index the extracted information in a text index;~~

~~a unit~~ the token processing module configured to determine structural connectivity information of each of the recognized chemical name fragments and ~~recognized~~ the substructures that do not contain a number, and to index representations of the recognized chemical name fragments and the ~~recognized~~ the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index;

~~a unit~~ the token processing module configured to store the text index in association with ~~indexed representations~~ the chemical substructure index ~~in a searchable index~~; ~~and~~

a ~~unit to provide~~ searcher module and a graphical user interface configured to search the ~~searchable~~ text index and the chemical substructure index, where the search comprises first entering one or more chemical fragment names and ~~entering~~ then selecting graphical representations of one or more substructures ~~in a representation form~~, ~~where the entering is by at least one of text form or graphical selection~~ where the selecting comprises using the graphical user interface as a pointer to a graphical list of substructures; and

the graphical user interface configured to receive a search result, where the search result is an intersection of the chemical substructure index and the text index, identifying at least one document where there are found chemical compounds that contain the selected substructures, and connectivity specified by the one or more chemical fragment names and the selected substructures.

20. (Currently Amended) ~~A~~ The system as in claim 19, wherein the ~~unit to extract is further to extract keywords from the text document and index the keywords in the text index, and wherein the~~ search further comprises ~~selecting a graphical representation of one or more substructures~~ first entering the one or more chemical fragment names and additionally entering at least one keyword, and where the search result is identifying at least one document where there are found the at least one keyword, the chemical compounds that contain the selected substructures, and the connectivity specified by the one or more chemical fragment names and the selected substructures.

21. – 24. (Cancelled)

25. (Currently Amended) ~~A~~ The system as in claim 19, where said ~~unit~~ token processing module that is configured to determine~~s~~ the structural connectivity information is further configured to look~~s~~ up recognized fragments and substructures in a structure dictionary.

26. (Currently Amended) ~~A~~ The system as in claim 19, where the token processing module configured to index representations is further configured to ~~comprise MOL type representations and SMILES type representations~~

test if each of the recognized chemical name fragments occur in a SMILES fragment dictionary, where if it does occur in the SMILES fragment dictionary the token processing module is configured to add the chemical name fragment to the chemical substructure index as the SMILES representation, and test if each of the recognized chemical name fragments occur in a MOL file fragment dictionary, where if it does occur in the MOL file dictionary the token processing module is configured to add the chemical name fragments to the chemical substructure index as the MOL file representation.

27. (Currently Amended) ~~A~~ The system as in claim 19, where said plurality of dictionaries ~~comprise a~~ consists of the dictionary of common chemical prefixes and ~~a~~ the dictionary of common chemical suffixes.

28. (Currently Amended) ~~A~~ The system as in claim 19, where said plurality of dictionaries ~~comprise~~ consists of the dictionary of common chemical prefixes, the dictionary of common chemical suffixes, and a dictionary of stop words to eliminate erroneous chemical name fragments.

29. (Currently Amended) ~~A~~ The system as in claim 19, further comprising ~~a unit~~ said token processing module is further configured to filter recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

30. (Currently Amended) ~~A~~ The system as in claim 19, where ~~chemical name fragments are~~ the tokenizer module is further ~~recognized~~ configured to recognize chemical name fragments by using common chemical word endings.

31. (Currently Amended) ~~A~~ The system as in claim 19, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed from between chemical name fragments as a function of context.

32. (Currently Amended) ~~A~~ The system as in claim 19, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

33. (Currently Amended) ~~A~~ The system as in claim 32, where the punctuation comprises at least one of a parenthesis, a square bracket, a hyphen, a colon and a semi-colon.

34. (Currently Amended) ~~A~~ The system as in claim 32, where the characters comprise ~~at least one of~~ upper case C, O, R, N and H.

35. (Currently Amended) ~~A~~ The system as in claim 32, where the characters comprise ~~strings of at least one of~~ lower case xy, ene, ine, yl, ane and oic.

36. (Currently Amended) ~~A~~ The system as in claim 19, further comprising an input tokenizer ~~unit~~ module configured to receive documents to be processed to provide a sequence of tokens.

37. – 41. (Cancelled)

42. (Currently Amended) ~~A computer program product as in claim 37, wherein the search further comprises entering at least one search term, and where a search results in an intersection of the indexed representations and the text index, to identify at least one document that contains a reference to a corresponding chemical compound~~ The system as in claim 43, where said plurality of dictionaries consists of the dictionary of common chemical prefixes, the dictionary of common chemical suffixes, and a dictionary of stop words to eliminate erroneous chemical name fragments.

43. (Currently Amended) A system comprising a plurality of computers at least two of which are coupled together through a data communications network, said system comprising:

a ~~unit~~ tokenizer and a token processing unit configured to parse text of a text document and assign semantic meaning to words of the parsed sentences, where assigning comprises applying a plurality of regular expressions, rules and dictionaries consisting of a common chemical prefix dictionary and a common chemical suffix dictionary to recognize chemical name fragments;

~~a unit~~ the token processing unit configured to recognize any substructures present in the chemical name fragments;

the token processing module configured to extract keywords associated with the recognized chemical name fragments and the substructures of the text document and to index the extracted keywords in a text index;

the token processing module configured to add each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

~~a unit to extract information associated with the recognized chemical name fragments and substructures from the text document and index the extracted information in a text index;~~

~~a unit~~ the token processing module configured to determine structural connectivity information of each of the recognized chemical name fragments and ~~recognized~~ the substructures that do not contain a number;

~~a unit~~ the token processing module configured to index representations of the recognized chemical name fragments and the ~~recognized~~ substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index;

~~a unit~~ the token processing module configured to store the text index in association with the ~~indexed representations in a searchable index~~ chemical substructure index; ~~and~~

a ~~unit to provide~~ searcher module and a graphical user interface configured to search the ~~searchable~~ text index and the chemical substructure index, where the search comprises first entering search terms comprising one or more chemical fragment names and then ~~entering~~

selecting graphical representations of one or more substructures ~~in a representation form~~, ~~where the entering is by at least one of text form or graphical selection~~ where the selecting comprises using the graphical user interface as a pointer to a graphical list of substructures; and

the graphical user interface configured to receive a search result, where the search result is an intersection of the chemical substructure index and the text index, identifying at least one document where there are found chemical compounds that contain a reference to the search terms and the one or more substructures.

44. (Currently Amended) ~~A~~ The system as in claim 43, wherein the ~~the~~ search further comprises ~~entering at least one search term, and wherein a search results in an intersection of a the indexed representations and the text index, to identify at least one document that contains a reference to a corresponding chemical compound~~ first entering the one or more chemical fragment names and additionally entering at least one keyword, and where the search result is identifying at least one document where there are found the at least one keyword, the chemical compounds that contain the selected substructures, and the connectivity specified by the one or more chemical fragment names and the selected substructures.

45. (Currently Amended) ~~A~~ The system as in claim 43, where said token processing module ~~unit that determines structural connectivity information~~ is further configured to ~~looks~~ look up recognized fragments and substructures in a structure dictionary.

46. (Currently Amended) ~~A~~ The system as in claim 43, where the indexing representations ~~comprise MOL type representations and SMILES type representations~~ of the recognized chemical name fragments and the substructures comprises:

testing if each of the recognized chemical name fragments occur in a SMILES fragment dictionary, where if it does occur in the SMILES fragment dictionary then adding the chemical name fragment to the chemical substructure index as the SMILES representation, and testing if each of the recognized chemical name fragments occur in a MOL file fragment dictionary, where if it does occur in the MOL file dictionary then adding the chemical name fragments to the chemical substructure index as the MOL file representation.